



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2018

---

## **Keystone taxa as drivers of microbiome structure and functioning**

Banerjee, Samiran ; Schlaeppli, Klaus ; van der Heijden, Marcel G A

**Abstract:** Microorganisms have a pivotal role in the functioning of ecosystems. Recent studies have shown that microbial communities harbour keystone taxa, which drive community composition and function irrespective of their abundance. In this Opinion article, we propose a definition of keystone taxa in microbial ecology and summarize over 200 microbial keystone taxa that have been identified in soil, plant and marine ecosystems, as well as in the human microbiome. We explore the importance of keystone taxa and keystone guilds for microbiome structure and functioning and discuss the factors that determine their distribution and activities.

DOI: <https://doi.org/10.1038/s41579-018-0024-1>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-168373>

Journal Article

Accepted Version

Originally published at:

Banerjee, Samiran; Schlaeppli, Klaus; van der Heijden, Marcel G A (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews. Microbiology*, 16(9):567-576.

DOI: <https://doi.org/10.1038/s41579-018-0024-1>

**Keystone taxa in microbiology: roles, drivers and importance for  
microbiome functioning**

Samiran Banerjee<sup>1\*</sup>, Klaus Schlaeppi<sup>1</sup>, Marcel G.A. van der Heijden<sup>1,2,3</sup>

<sup>1</sup>Department of Agroecology and Environment, Agroscope, Zurich 8046, Switzerland

<sup>2</sup>Department of Evolutionary Biology and Environmental Studies, University of Zürich,  
Zürich 8057, Switzerland

<sup>3</sup>Institute of Environmental Biology, Faculty of Science, Utrecht University, 3508 TC,  
Utrecht, The Netherlands

\* Corresponding author: +41 58 462 4406; [samiran.banerjee@agroscope.admin.ch](mailto:samiran.banerjee@agroscope.admin.ch)

## **Abstract**

Microbes play a key role in the functioning of ecosystems. Recent studies have shown that

microbial communities harbor keystone taxa, taxa that have a significant influence on

community composition and microbiome performance irrespective of their abundance. Here

5 we propose a definition of keystone taxa in microbiology and summarize reports of such taxa

from soil, plant and marine ecosystems as well as the human microbiome. We illustrate the

importance of keystone taxa and keystone guilds for ecosystem services, and discuss the

factors that determine their distribution and activities. This article highlights the relevance of

keystone taxa as the drivers of microbiome structure and functioning.

10

15

20

25

## Introduction

The role of microbial communities in ecosystem functioning is unequivocal<sup>1,2</sup>. Microbes are key drivers of a wide range of ecosystem services including soil nutrient cycling, plant growth promotion, marine biogeochemical processes, and maintenance of human health<sup>3–6</sup>.

30 In recent years, microbial networks have been used to visualize co-occurrence among members in communities<sup>3,7–10</sup>. Microbial networks enable testing of ecological theories, assessment of which was once postulated to be a major impediment in microbial ecology<sup>11,12</sup>. The concept of co-occurrence and network thinking in ecology was proposed in 2005<sup>13</sup> and since then, microbial ecologists have shown particular interest in network analysis<sup>7,14–19</sup>,  
35 resulting in a large body of studies demonstrating microbial co-occurrence patterns in a diverse range of soil<sup>7</sup>, plant<sup>20</sup>, marine<sup>21</sup> ecosystems and in the human microbiome<sup>22,23</sup> (**BOX 1**). Reports are also available from the Antarctic<sup>24</sup> and Arctic ecosystems<sup>25,26</sup>. Amidst other salient features, microbial networks can statistically identify *keystone taxa*<sup>27</sup>.

The tenet of keystone taxa was originally proposed by ecologist Robert T. Paine in  
40 1966. In a classic experiment, he demonstrated that the removal of sea stars (*Pisaster ochraceus*), a common predator of mussels, had a dramatic impact on the shoreline ecosystem community and local biodiversity at Makah Bay, Washington<sup>28</sup>. Other examples of keystone taxa include the Canadian beaver and African elephant in the animal kingdom, leguminous *Trifolium* in the plant kingdom, and *Porphyromonas gingivitis* and *Bacteroidetes*  
45 *thetaiotaomicron* in the human microbiome<sup>29–32</sup>. Keystone taxa have been frequently referred to as ecosystem engineers due to their large influence in the community<sup>33</sup>. In microbial communities, examples of such taxa are now available from a diverse range of environments<sup>24–26,31,34–59</sup> and their reports are continuously increasing (**BOX 1**).

## 50 Definition of keystone taxa in microbial communities

The definition of keystone taxa has garnered different lines of thought <sup>60–62</sup>. The definition proposed by Paine (1969) mainly suggests that keystone taxa are important for community structure and integrity, and their influence is nonredundant <sup>63</sup>. Later, Power et al. (1996) defined keystone taxa by introducing the concept of “community importance”, calculated from proportional biomass and traits. Cotten-Jones and Whitaker (2012) presented the evolution of the term “keystone taxa” in ecology and how its overuse and misuse (e.g., keystone mutualist, keystone modifier, reverse keystone, etc.) have resulted in considerable confusion about the actual meaning. Readers are referred to their critical appraisal for further information on keystone taxa in ecology. Thus, there is no uniformly accepted operational definition of keystone taxa in ecology, especially in microbial ecology. Keystone taxa such as *Bacteroides thetaiotaomicron*, an anaerobic symbiont found in the human intestine, do not fulfil the low-abundance criterion <sup>64</sup>. However, in view of their impact, they are considered as keystone taxa. The definition of keystone taxa in the microbial world should also take into account spatiotemporal considerations, due to rapid microbial turnover in both time and space. Thus, we propose the following definition: *keystone taxa are the taxa which have major influence on microbiome composition and function at a particular space or time. These taxa often, but not always, have an over-proportional influence in the community, relative to their abundance.*

## **Microbial networks and keystone taxa**

With the advent of next generation sequencing, millions of sequences are now available from various environments. Network analysis can disentangle microbial co-abundance and empower microbial ecologists to gain a comprehensive insight into the microbial community structure and assembly patterns <sup>3,65</sup>. Several algorithms are available to construct microbial

networks and these algorithms have already been reviewed previously<sup>19,65,66</sup> and thus, for brevity, this article will only present an overview (**BOX 1**). Perhaps one of the most useful features of network analysis is that it can identify ‘hubs’ or keystone operational taxonomic units (OTUs) that are highly associated in a microbiome (**Figure 1**). Unlike random networks with a Poisson distribution, scale-free or small-world networks with a power-law distribution comprise such hubs or highly connected nodes (reviewed in REFS<sup>67,68</sup>). These hubs have been proposed repeatedly as keystone taxa as their removal has been computationally shown to cause a shift in the composition and functioning of a microbiome<sup>70,71</sup>. While high betweenness centrality was previously used to identify keystone taxa statistically in several studies<sup>16,39,40</sup>, Berry and Widder (2014) recently showed that high mean degree, high closeness centrality and low betweenness centrality can be collectively used to identify keystones with 85% accuracy<sup>27</sup>. Subsequently, these scores have been used to find putative keystone taxa in microbial networks in recent studies<sup>12,41,42</sup>. In their appraisal, Cottee-Jones and Whitaker (2012) highlighted the importance of a *quantifiable threshold* for consistent identification and validation of keystone taxa. We recommend that the combined score proposed by Berry and Widder (2014) serves as the threshold for defining keystone taxa in microbiology.

## Recent evidence of keystone taxa

### *Computational inference*

Numerous studies have used network based scores to identify putative keystone taxa in various environments (**Table 1**; **Table S1** in *Supporting Information*). Zhou *et al.* (2011) identified hubs in microbial networks in grassland soils while Lupatini *et al.* (2014) found that the Pampa and Cerrado biomes in Brazil harbored different keystone taxa. Ma *et al.* (2016) conducted a continental scale network analysis and showed that bacterial keystones

belonged to *Alphaproteobacteria* and *Actinobacteria* and fungal keystones belonged to *Pezizomycotina*. Keystone taxa have also been identified in Arctic<sup>25,26,44,46</sup> and Antarctic ecosystems<sup>24</sup>, which were not numerically dominant in the communities. Similar reports are also available for microbial communities in contaminated soils<sup>47,48</sup> and aquatic systems<sup>24,55,72</sup>. Interestingly, our literature review revealed that various members of *Rhizobiales* and *Burkholderiales* were identified as keystone taxa in different studies and across different ecosystems (Table 1; Table S1). The order *Rhizobiales* not only comprises nitrogen fixers including members of *Rhizobium* and *Bradyrhizobium*, but also other important groups such as *Methylobacterium* that are abundant in phyllosphere but also known to be endosymbiotic<sup>73</sup>. On the other hand, *Burkholderiales* includes important genera such as *Bordetella*, *Ralstonia*, *Oxalobacter* that are well-known pathogens, but also one of the most versatile and diverse terrestrial groups, *Burkholderia*. This observation does not mean that all *Rhizobiales* and *Burkholderiales* can be considered keystone taxa (e.g., many taxa in those clades are subordinate taxa in microbial communities and have no major influence on community composition or functioning). Computational inference of *Rhizobiales* and *Burkholderiales* as keystones can also be due to their sheer abundance in various environments. However, our review indicates that the likelihood of finding a keystone taxon within these two clades is higher. Further targeted research on their role as keystones in influencing microbial functioning is needed.

### *Empirical evidence*

Empirical evidence of keystone taxa is mainly available from the human microbiota, where keystones have been linked to a range of processes including inflammation, colon and gastric cancer, starch degradation and stabilization of the microbiome<sup>22,23,31,32,56–59,64</sup> (see **Table S1** for details). Perhaps one of the most prominent keystones in humans is *Bacteroides fragilis*,

which spurred the “alpha-bug” or “keystone pathogen” hypothesis <sup>58,64</sup>. Other examples of keystone taxa include *Porphyromonas gingivitis* <sup>32</sup>, *Bacteriodes thetaiotaomicron* <sup>31</sup>, *Ruminococcus bromii* <sup>59</sup>, *Methanobrevibacter* <sup>64</sup> and *Helicobacter pylori* <sup>56</sup>. These taxa can cause significant shifts in the composition and functioning of the oral and gut microbiome.

130 Examples are also available from other microbiomes where keystones have been identified through network based scores and linked to microbiome functioning and ecosystem processes. Agler *et al.* (2016) found that the effects of abiotic factors and host genotypes on microbial communities are mediated via microbial keystone taxa. This not only supports the relevance of keystone taxa but also provides evidence of their importance for plant  
135 microbiome functioning. Nitrogen-fixing rhizobia have been proposed as keystone taxa and their abundance has been shown to significantly improve plant productivity and community evenness <sup>74</sup>. Banerjee *et al.*, (2016b) demonstrated keystone taxa in bacterial and fungal communities during organic matter decomposition in an agricultural soil and, interestingly, these taxa were also identified as keystones for organic matter transformation in agricultural  
140 soil in a recent study <sup>41</sup>, indicating the importance of similar keystones for specific habitats and processes. Recently, Herren and McMahon (2017) found that low-abundant keystone taxa which are highly connected in the microbiome can explain microbiome compositional turnover better than all taxa combined. Indeed, such encouraging reports highlight the relevance of keystone taxa for microbiome composition and functioning.

## 145 **Challenges in identifying keystone taxa**

### *Correlation ain't causation*

Keystone taxa identified using network based scores were linked to ecological processes in many studies (e.g., see **Table S1**), indicating the suitability of this method. However, to  
150 obtain convincing evidence, network-based scores need to be complemented with



experimental evidence on the impact of the keystone taxa on microbiome composition and function. The detection of keystone taxa using network-based scores can be biased by habitat filtering, and networks can show positive associations between non-interacting members in microbial communities. Moreover, network scores and co-occurrence patterns are ultimately based on correlations and they must be interpreted with caution as *correlation does not mean causation*. Statistical analyses such as structural equation modelling (SEM) can be employed to move beyond correlation analysis and explore causal relationships among keystone taxa and microbiome composition or function. SEM is an advanced multivariate statistical approach that identifies such causal relationships and generates strong and distinct links between theoretical and experimental ideas<sup>76</sup>. The strength of SEM lies in the fact that it is theory oriented and not null hypothesis based and thus, it provides a framework to interpret complex networks involving numerous response and predictor variables. Upon assessing the univariate and multivariate normality, an initial model is generated based on the existing knowledge, site information and background data<sup>77</sup>. Subsequently, a  $\chi^2$  test is conducted to assess if the covariance structure indicated by the model adequately fits the covariance structures, where a non-significant  $\chi^2$  test suggests sufficient model fit. Importantly, the requirement of a minimum sample size of 50 must be followed while employing SEM<sup>76</sup>. Ascertaining disproportionality of influence for keystones can also be a challenge<sup>78</sup>. A recent study used sparse linear regression with bootstrap aggregation in a discrete-time Lotka-Volterra model to identify *Bacteriodes fragilis* and *Bacteriodes stercosis* as the keystone taxa with disproportionate influence on the gut microbiome structure<sup>22</sup>. This is encouraging because the algorithm statistically identified *Bacteriodes fragilis* as a keystone taxon, which already has empirical validity<sup>64</sup>.

*Characterization and manipulation*

While experimental manipulation (e.g., removing a putative keystone to assess the impact) is the popular choice among plant and animal ecologists, one of the fundamental challenges that microbiologists are confronted with is the characterization and manipulation of such taxa.

Manipulating growth or co-culturing microbes on nutrient media or petri-dishes or in

microcosms can be challenging owing to individual physiological requirements. There have been some exciting strategies in the last few years that employed novel approaches to tackle the uncultivability issue. Nichols *et al.* (2010), for example, developed the *ichip* that employs the diffusion chamber method and allows *in situ* cultivation of novel microbiota. Similarly, a “microbial trap” has been developed to capture and culture *Actinobacteria* in *in situ*

conditions<sup>80</sup>. On the other hand, *on-chip* microbial culture coupled with surface plasmon resonance allows *in situ* detection of novel and rare microbes<sup>81</sup>. Droplet-based microfluidic technology also offers the opportunity to mimic natural conditions and co-cultivate synergistic microbial communities<sup>82</sup>, while the *microbiome-on-a-chip* approach enables the study of microbial networks and their associations with host plants<sup>83</sup>. Future studies may

wish to include such promising approaches to isolate and characterize keystone taxa from various ecosystems. Removal of keystone taxa may lead to an alternative stable state (*sensu*<sup>84</sup>) of the microbial network, which results in dysfunction or even renewed functioning if the keystone originally had a negative impact. Future studies may also wish to experimentally manipulate microbial network structure in synthetic communities to assess whether the removal of keystone taxa disrupts microbiome functioning.

### **Keystone taxa and influence on microbiome**

The influence irrespective of their abundance distinguishes keystone taxa from dominant taxa. A dominant species often affects ecosystem functioning or a specific process

exclusively by virtue of sheer abundance (**Figure 2A**), whereas keystone taxa might exert its

influence on microbiome functioning irrespective of its abundance. The importance of keystone taxa may also be related to the *broadness of a process*, i.e., a process involving many steps and operated by functionally and taxonomically distinct microbial groups<sup>1,85</sup>. For example, dominant taxa with large biomass or major energy transformations might influence broad processes such as denitrification or organic matter decomposition. On the other hand, the influence of rare keystone taxa might be stronger if a process is “narrow”, consisting of single steps (e.g., nitrogen fixation, ammonia oxidation, etc.) and operated by a small group of microbes<sup>1,85</sup>. We postulate that the influence of rare keystone taxa on an ecosystem process is inversely proportional to the broadness. However, it should be noted that some keystone taxa such as *Bacterioidetes thetaiotaomicron* in the human intestine can be numerically dominant as well and thus the distinction between dominant and less abundant keystone taxa is not always true. Thus, whether numerically inconspicuous keystone taxa are more influential on narrow processes is a hypothesis that needs further investigation.

Keystone taxa might resort to a range of strategies to exert an influence on a microbiome. For example, they might function via *intermediate* or *effector groups*. These are the groups whose abundance can be selectively modulated to regulate community structure and functioning<sup>23,64</sup>. Such selective modulation might include promotion (commensalism) or suppression (ammensalism) of effector groups by secreting metabolites, antibiotics or toxins with themselves being unaffected. In humans, *Porphyromonas gingivitis* exerts its influence by causing dysbiosis, the community-wide change in relative abundance of other microbes, which thereby results in inflammation<sup>32</sup>. Similarly, certain strains of *Pseudomonas fluorescens* produce a secondary metabolite (2,4-diacetylphloroglucinol) that suppresses *Gaeumannomyces graminis* var *tritici*, which causes the take-all disease in wheat<sup>86</sup>.

Alternatively, keystone taxa might produce bacteriocins to selectively alter microbiota composition. For example, bacteriocin production by *Enterococcus faecalis* can induce niche

competition in the gastrointestinal tract to change microbiota composition<sup>87</sup>. keystones might also engage in synergistic relationships and change the abundance of their partners. Some members of *Burkholderia* can act as an endosymbiont in arbuscular mycorrhizal fungi to change the abundance and community characteristics of this important group in soil<sup>88</sup>, which subsequently may alter plant community richness and productivity<sup>89</sup>. Thus, keystone taxa can employ different strategies to shape the microbiota in their favor, but the selection of a particular strategy would depend on the microenvironment.

### Putative drivers of keystone taxa

The presence of keystone taxa in a microbiome does not necessarily guarantee their influence because a number of factors may still determine their distribution and efficacy (**Figure 2B**). For example, *spatiotemporal heterogeneity* can be a major driver of the abundance and distribution of keystone taxa<sup>60,62,78</sup>. This is particularly true for soil, which is one of the most heterogeneous and multifaceted environments. Similarly, seasonal variability determines the structural and compositional properties of microbiomes in an environment, and as such, a keystone might only be present in a specific season or time period.

The occurrence and functioning of a keystone will also depend on its *position in the microbiome*. Recently, the tenet of *core microbiomes* and *holobionts* has been proposed for humans<sup>6,90</sup>, plants<sup>91</sup> and readers are referred to Hamady and Knight (2009) and

Vandenkoornhuyse et al. (2015) for the taxonomic and functional definitions of a core microbiome. Keystone taxa might belong to the core microbiome that is consistently present in an environment regardless of changes in environmental conditions<sup>90,91</sup>. In a seminal paper, Turnbaugh et al (2009) first presented the evidence of a core gut microbiome in obese and lean twins. Recently, Yeoh et al. (2017) found an evolutionarily conserved core microbiome in plant roots and it is also fascinating that some of the well-known keystone taxa such as

*Rhizobium*, *Bradyrhizobium* and *Burkholderia* are also part of the core root microbiome. The contribution of keystone taxa would be higher if they are part of the core microbiome, highlighting the importance of such taxa for microbiome functioning<sup>23</sup>.

Microbiomes can also harbor *keystone guilds* or groups of keystone taxa with similar functioning<sup>62</sup> (**Figure 2C**). Examples of such guilds that can alter structure and dynamics of ecosystems are common in the animal world<sup>94</sup>. Perhaps the most famous example is the three species of kangaroo rats that are a keystone guild in the Chihuahuan desert and have a strong impact on local biodiversity and biogeochemical processes<sup>94</sup>. In the microbial world, keystone guilds may arise based on a number of factors such as complementary resource acquiring strategies, resource sharing, niche partitioning, spatiotemporal coherence, etc.<sup>29,78</sup>. While numerically inconspicuous keystone taxa might have a greater influence on narrow processes, a keystone guild consisting of diverse keystone taxa within a community might also influence a broad process. Jones et al. (2014), for example, showed that certain key guilds of co-occurring denitrifiers can play a significant role in denitrification, a broad process operated by heterogeneous groups of microbes. We expect that examples of such keystone guilds will continue to rise in the future. Indeed, such guilds may be particularly powerful if they belong to the core microbiome.

Keystone taxa or members of keystone guilds might be *functionally redundant* or their effect might be *context dependent*. For example, the effect might only be observed under a specific scenario whereas in other scenarios the species might be non-keystone or even nonexistent. Such context dependency or conditionality may be more common in environments with turbulence or high spatiotemporal variability<sup>62</sup>. Thus, keystone taxa might not be always keystones in environments. A plausible challenge for assessing keystone taxa is also the fact that there might be a *hysteresis effect* i.e., time lag between the change in keystone taxa and their influence on microbiome functioning. With rapid microbial turnover,

identifying such lags can be a daunting task. Nonetheless, while we provided a list of potential drivers of keystone taxa, this is not exhaustive as there may be other factors influencing these taxa in the microbial world.

## 280    **Keystone in the light of rare species concept**

Keystone taxa underline the importance of numerically inconspicuous taxa for microbiome functioning, which is also congruent with the rare taxa concept. Indeed, the fundamental premise of keystone and rare taxa is the same: species abundance is not the best determinant of its contribution to the community<sup>96</sup>. Importance of rare microbes have been observed for  
285    many biogeochemical processes including nitrification, denitrification, methanogenesis, methanotrophy and sulfate reduction (reviewed in REFS<sup>96,97</sup>). A seminal example of the significance of rare microbes came from the study by Pester *et al.* (2010), who showed that *Desulfosporosinus* sp. ,which only represents 0.06% of the total community, plays a pivotal role in sulfate reduction and carbon flow in peatland soils. Rare biosphere has also been  
290    found important in human microbiome and even in depauperate ecosystems<sup>97</sup>. Evidence of such low abundant taxa with an overproportional influence obviously raises the possibility that members of rare biosphere can also be keystone taxa.

## **Next frontier**

### 295    *Unexplored areas*

Network scores have been popular to statistically identify keystone taxa in recent years and it is important to continue this momentum to strengthen keystone repertoire. For example, information on keystone taxa from desert, tropical forest or vadose zone are rare or not available yet. Similarly, knowledge about how keystones respond to environmental  
300    disturbance, pathogen attack in plants, or medical intervention in humans would be valuable.

For example, it could be tested whether keystone taxa help microbiome resilience against perturbations. Keystone taxa in plant invasion is an equally interesting area, especially in the light of observations that some invasive tree species cannot establish without their microbial symbionts or that invasive species alter the soil microbiome<sup>99</sup>. Moreover, our knowledge of  
305 fungal, archaeal and protistan keystones is negligible and only a few studies have considered fungal-bacteria, or fungal-archaeal and bacterial-archaeal co-occurrence networks<sup>38,39</sup>. Thus, a cross-domain network may reveal how members of different taxonomic groups associate with each other or if they have overlap in resource sharing. An intriguing question is whether keystones in microbial communities follow similar ecological principles (e.g., drift, dispersal,  
310 diversification, environmental selection; *sensu*<sup>12</sup>) as keystones in plant or animal kingdoms.

#### *Linking keystone taxa to ecosystem processes*

Linking community structure to function is a central goal in microbial ecology<sup>11</sup> and it is necessary to extend microbial co-occurrence patterns and keystone taxa to ecosystem  
315 processes (**Figure 3**). Studies investigating keystone taxa may wish to include promising culturing approaches to explore complex ecological relationships such as commensalism and ammensalism in natural conditions and assess the effect of keystones. The actual importance of keystone taxa to microbiome functioning and ecosystem processes can only be derived from robust functional profiling using the latest tools such as RNA-stable isotope probing<sup>100</sup>  
320 coupled with metatranscriptomics or metaproteomics. Upon identifying keystone taxa in an environment, it could also be tested if there are *structural keystones* and *functional keystones* depending on whether they affect microbiome structure or functioning. Since any change in microbiome structure may also have consequences for microbiome functioning, a clear distinction between structural and functional keystones in microbial communities is  
325 questionable. Nonetheless, the latest molecular tools have empowered microbiologists to test

such theories and ideas. The contribution of microbial communities for ecosystem processes is often missing or insignificant in ecosystem models <sup>5</sup>. These models mostly consider the overall community characteristics (abundance, composition, diversity), which might blur the actual contribution of important microbial members. Keystone taxa observed across habitats and studies might be the missing piece of the puzzle that could help microbial ecologists *explain the unexplained variation* in ecosystem processes.

### **Concluding remarks**

Beyond the dominant taxa with large biomass or major energy transformations, keystone taxa can orchestrate microbiota to perform ecosystem processes. The number of studies investigating keystones is rising exponentially and we expect this to continue in the next few years. With the noticeable exception of keystone pathogens in the human microbiome, keystone taxa are mainly identified through network based scores. To quantify the contribution of keystone taxa to microbiome functioning, microbial ecologists need to employ the latest microbiological and analytical techniques, and move beyond mere description of their occurrence in different ecosystems. The conceptual and empirical insights presented in this opinion article can be useful for future microbiological studies.

### **Acknowledgements**

We thank the Swiss National Science Foundation (Grant No. 31003A\_166079) for the financial support. We also thank three reviewers whose constructive comments and insightful suggestions greatly improved the quality of the manuscript.



## References

1. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* (2017). doi:10.1038/nrmicro.2017.87
2. Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505–511 (2014).
3. Fuhrman, J. A. Microbial community structure and its functional implications. *Nature* **459**, 193–9 (2009).
4. van der Heijden, M. G. A., Bardgett, R. D. & Van Straalen, N. M. The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* **11**, 296–310 (2008).
5. Graham, E. B. *et al.* Microbes as engines of ecosystem function: When does community structure enhance predictions of ecosystem processes? *Front. Microbiol.* **7**, 1–10 (2016).
6. Hamady, M. & Knight, R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res.* **19**, 1141–1152 (2009).
7. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351 (2012).
8. Banerjee, S. *et al.* Determinants of bacterial communities in Canadian agroforestry systems. *Environ. Microbiol.* **18**, 1805–1816 (2016).
9. Zhang, Z. *et al.* Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J.* **8**, 881–893 (2014).
10. Fuhrman, J. A., Cram, J. A. & Needham, D. M. Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* **13**, 133–146 (2015).
11. Prosser, J. I. *et al.* The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* **5**, 384–392 (2007).
12. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M. & Relman, D. a. The Application of Ecological Theory. **336**, 1255–1262 (2012).
13. Proulx, S. R., Promislow, D. E. L. & Phillips, P. C. Network thinking in ecology and evolution. *Trends Ecol. Evol.* **20**, 345–353 (2005).
14. Zhou, J. *et al.* Functional molecular ecological networks. *MBio* **1**, e00169-10 (2010).
15. Reshef, D. N. *et al.* Detecting novel associations in large data sets. *Science (80-. )*. **334**, 1518–1524 (2011).
16. Ruan, Q. *et al.* Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **22**, 2532–2538 (2006).
17. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput. Biol.* **8**, 1–11 (2012).
18. Faust, K. *et al.* Microbial co-occurrence relationships in the Human Microbiome. *PLoS Comput. Biol.* **8**, (2012).
19. Weiss, S. *et al.* Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J* **10**, 1–13 (2016).
20. Agler, M. T. *et al.* Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLoS Biol.* **14**, 1–31 (2016).
21. Gilbert, J. a *et al.* Defining seasonal marine microbial community dynamics. *ISME J.* **6**, 298–308 (2012).
22. Fisher, C. K. & Mehta, P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. *PLoS One* **9**, 1–10 (2014).
23. Shetty, S. A., Hugenholtz, F., Lahti, L., Smidt, H. & de Vos, W. M. Intestinal microbiome landscaping: Insight in community assemblage and implications for

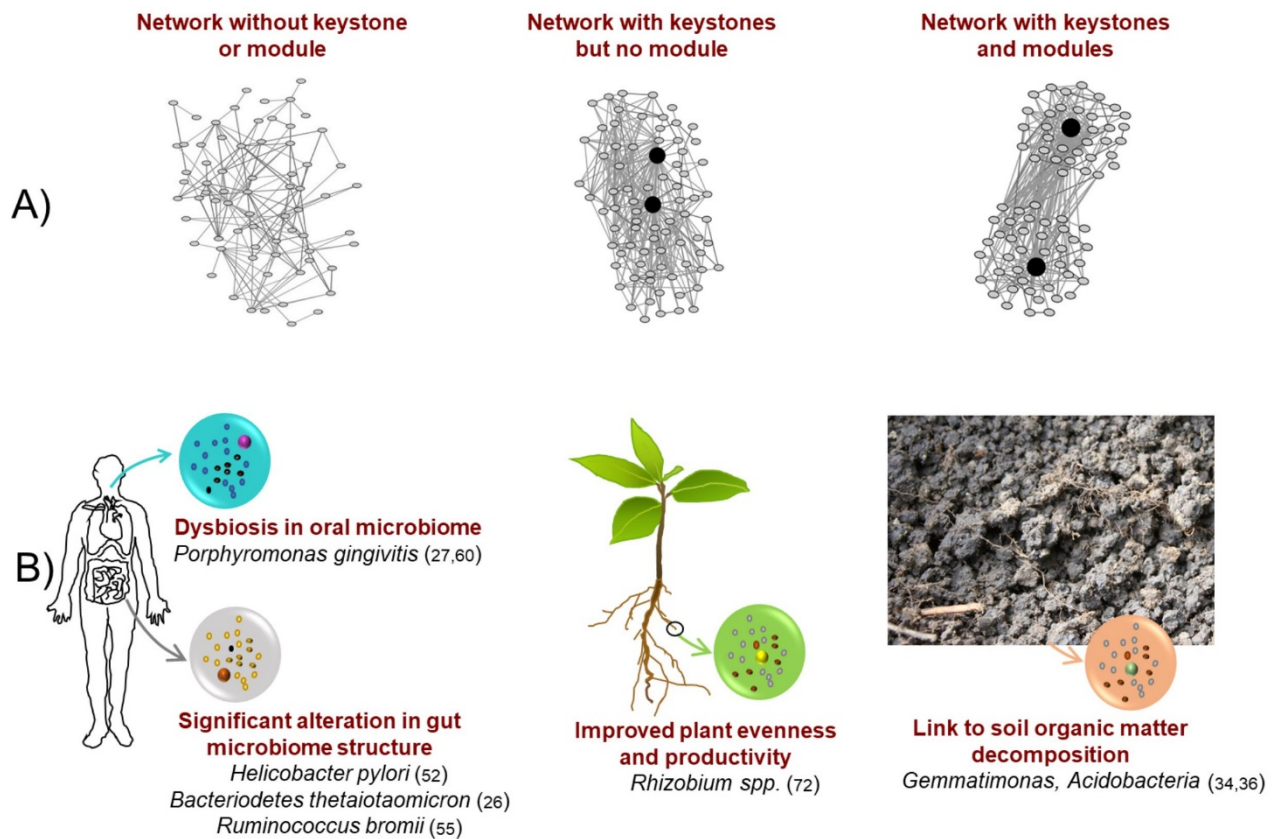
- 400 microbial modulation strategies. *FEMS Microbiol. Rev.* **41**, 182–199 (2017).
24. Vick-Majors, T. J., Priscu, J. C. & Amaral-Zettler, L. A. Modular community structure suggests metabolic plasticity during the transition to polar night in ice-covered Antarctic lakes. *ISME J.* **8**, 778–89 (2014).
25. Gokul, J. K. *et al.* Taxon interactions control the distributions of cryoconite bacteria colonizing a High Arctic ice cap. *Mol. Ecol.* **25**, 3752–3767 (2016).
- 405 26. Comte, J., Lovejoy, C., Crevecoeur, S. & Vincent, W. F. Co-occurrence patterns in aquatic bacterial communities across changing permafrost landscapes. *Biogeosciences* **13**, 175–190 (2016).
27. Berry, D. & Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **5**, 1–14 (2014).
- 410 28. Paine, R. T. Food Web Complexity and Species Diversity. *Am. Nat.* **100**, 65–75 (1966).
29. Nunez, M.A., Dimarco, R. D. Keystone Species. *The Berkshire Encyclopedia of Sustainability: Ecosystem Management and Sustainability* (2012).
- 415 30. Hector, A. *et al.* Plant diversity and productivity experiments in European grasslands. *Science* (80-. ). **286**, 1123–1127 (1999).
31. Curtis, M. M. *et al.* The gut commensal bacteroides thetaiotaomicron exacerbates enteric infection through modification of the metabolic landscape. *Cell Host Microbe* **16**, 759–769 (2014).
- 420 32. Hajishengallis, G. *et al.* Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. *Cell Host Microbe* **10**, 497–506 (2011).
33. Dunne, J. A., Williams, R. J. & Martinez, N. D. Network structure and biodiversity loss in food webs: robustness increase with connectance. *Ecol. Lett.* **5**, 558–567 (2002).
- 425 34. Deng, Y. *et al.* Molecular ecological network analyses. *BMC Bioinformatics* **13**, 113 (2012).
35. Lupatini, M. *et al.* Network topology reveals high connectance levels and few key microbial genera within soils. *Front. Environ. Sci.* **2**, 1–11 (2014).
- 430 36. Zhou, J., Deng, Y., Luo, F., He, Z. & Yang, Y. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO<sub>2</sub>. *MBio* **2**, e00122-11 (2011).
37. Eldridge, D. J. *et al.* Soil-foraging animals alter the composition and co-occurrence of microbial communities in a desert shrubland. *ISME J.* **9**, 1–11 (2015).
- 435 38. Ma, B. *et al.* Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* **10**, 1–11 (2016).
39. Banerjee, S. *et al.* Network analysis reveals functional redundancy and keystone taxa amongst bacterial and fungal communities during organic matter decomposition in an arable soil. *Soil Biol. Biochem.* **97**, 188–198 (2016).
- 440 40. Jiang, Y. *et al.* Plant cultivars imprint the rhizosphere bacterial community composition and association networks. *Soil Biol. Biochem.* **109**, 145–155 (2017).
41. Li, F., Chen, L., Zhang, J., Yin, J. & Huang, S. Bacterial Community Structure after Long-term Organic and Inorganic Fertilization Reveals Important Associations between Soil Nutrients and Specific Taxa Involved in Nutrient Transformations. *Front. Microbiol.* **8**, (2017).
- 445 42. Liang, Y. *et al.* Long-term oil contamination alters the molecular ecological networks of soil microbial functional genes. *Front. Microbiol.* **7**, 1–13 (2016).
43. Wang, H. *et al.* Combined use of network inference tools identifies ecologically meaningful bacterial associations in a paddy soil. *Soil Biol. Biochem.* **105**, 227–235

- (2017).
44. Hill, R. *et al.* Temporal and spatial influences incur reconfiguration of Arctic heathland soil bacterial community structure. *Environ. Microbiol.* **18**, 1942–1953 (2016).
  45. Li, B. *et al.* Metagenomic and network analysis reveal wide distribution and co-occurrence of environmental antibiotic resistance genes. *ISME J.* **9**, 1–13 (2015).
  46. Yang, S. *et al.* Hydrocarbon degraders establish at the costs of microbial richness, abundance and keystone taxa after crude oil contamination in permafrost environments. *Sci. Rep.* **6**, 37473 (2016).
  47. Chao, Y. *et al.* Structure, variation, and co-occurrence of soil microbial communities in abandoned sites of a rare earth elements mine. *Environ. Sci. Technol.* **50**, 11481–11490 (2016).
  48. Jiao, S. *et al.* Bacterial communities in oil contaminated soils: Biogeography and co-occurrence patterns. *Soil Biol. Biochem.* **98**, 64–73 (2016).
  49. Shi, S. *et al.* The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. *Ecol. Lett.* **19**, 926–936 (2016).
  50. Yan, Y., Kuramae, E. E., De Hollander, M., Klinkhamer, P. G. & Van Veen, J. A. Functional traits dominate the diversity-related selection of bacterial communities in the rhizosphere. *ISME J.* **11**, 1–11 (2016).
  51. Geng, H., Tran-Gyamfi, M. B., Lane, T. W., Sale, K. L. & Yu, E. T. Changes in the structure of the microbial community associated with *Nannochloropsis salina* following treatments with antibiotics and bioactive compounds. *Front. Microbiol.* **7**, 1–13 (2016).
  52. Graham, E. B. *et al.* Deterministic influences exceed dispersal effects on hydrologically-connected microbiomes. *Environ. Microbiol.* 1–44 (2017). doi:10.1111/1462-2920.13720
  53. Ji, Y. *et al.* Structure and function of methanogenic microbial communities in sediments of Amazonian lakes with different water types. *Environ. Microbiol.* **18**, 5082–5100 (2016).
  54. Musat, N. *et al.* A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 17861–6 (2008).
  55. Zhao, D. *et al.* Network analysis reveals seasonal variation of co-occurrence correlations between Cyanobacteria and other bacterioplankton. *Sci. Total Environ.* **573**, 817–825 (2016).
  56. Maldonado-Contreras, A. *et al.* Structure of the human gastric bacterial community in relation to *Helicobacter pylori* status. *ISME J.* **5**, 574–579 (2011).
  57. Trosvik, P. & de Muinck, E. J. Ecology of bacteria in the human gastrointestinal tract—identification of keystone and foundation taxa. *Microbiome* **3**, 44 (2015).
  58. Wu, S. *et al.* A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nat. Med.* **15**, 1016–1022 (2009).
  59. Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* **6**, 1535–1543 (2012).
  60. Mills, L. S. & Doak, D. F. The Keystone-Species Concept in Ecology and Conservation. *BioScience* **43**, 219–224 (1993).
  61. Cottee-Jones, H. E. W. & Whittaker, R. J. The keystone species concept: a critical appraisal. *Front. Biogeogr.* **4**, 217–220 (2012).
  62. Power, M. E. *et al.* Challenges in the Quest for Keystones. *Bioscience* **46**, 609–620 (1996).
  63. Paine, R. T. A note on trophic complexity and community stability. *The American*

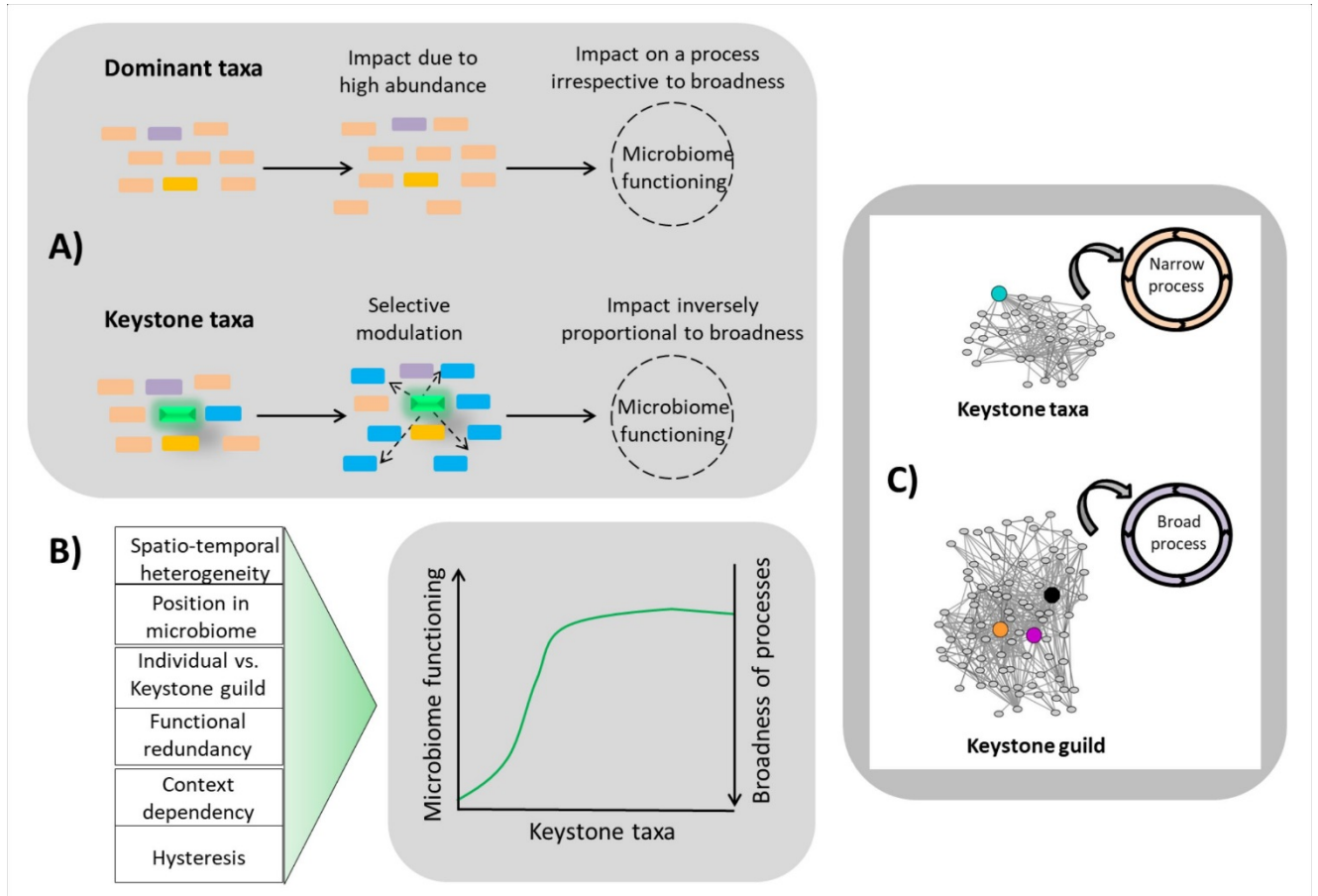
- 500 *Naturalist* **103**, 91–93 (1969).
64. Hajishengallis, G., Darveau, R. P. & Curtis, M. A. The keystone-pathogen hypothesis. *Nat. Rev. Microbiol.* **10**, 717–25 (2012).
65. Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **10**, 538–550 (2012).
- 505 66. Cardona, C., Weisenhorn, P., Henry, C. & Gilbert, J. A. Network-based metabolic analysis and microbial community modeling. *Curr. Opin. Microbiol.* **31**, 124–131 (2016).
67. Strogatz, S. H. Exploring complex networks. *Nature* **410**, 268–276 (2001).
68. Newman, M. E. J. The structure and function of complex networks. *Soc. Ind. Appl. Math. Rev.* **45**, 167–256 (2003).
- 510 69. Newman, M. E. J. The structure and function of complex networks. *E-Print Cond-Mat/0303516* **45**, 167–256 (2003).
70. van der Heijden, M. G. A. & Hartmann, M. Networking in the Plant Microbiome. *PLoS Biol.* **14**, 1–9 (2016).
- 515 71. Barabási, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
72. Steele, J. A. *et al.* Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* **5**, 1414–25 (2011).
73. Andreote, F. D. *et al.* Culture-independent assessment of rhizobiales-related alphaproteobacteria and the diversity of *Methylobacterium* in the rhizosphere and rhizoplane of transgenic eucalyptus. *Microb. Ecol.* **57**, 82–93 (2009).
- 520 74. Van Der Heijden, M. G. A. *et al.* Symbiotic bacteria as a determinant of plant community structure and plant productivity in dune grassland. *FEMS Microbiol. Ecol.* **56**, 178–187 (2006).
- 525 75. Herren, C. M. & McMahon, K. D. Small subsets of highly connected taxa predict compositional change in microbial communities. *bioRxiv* (2017). doi:<http://dx.doi.org/10.1101/159087>
76. Grace, J. B., Anderson, T., Olff, H. & Scheiner, S. On the specification of structural equation models for ecological systems. *Ecol. Monogr.* **80**, 67–87 (2010).
- 530 77. Lamb, E. G., Shirliff, S. J. & May, W. E. Structural equation modeling in the plant sciences: An example using yield components in oat. *Can. J. Plant Sci.* **91**, 603–619 (2011).
78. Mouquet, N., Gravel, D., Massol, F. & Calcagno, V. Extending the concept of keystone species to communities and ecosystems. *Ecol. Lett.* **16**, 1–8 (2013).
- 535 79. Nichols, D. *et al.* Use of ichip for high-throughput in situ cultivation of ‘uncultivable microbial species’. *Appl. Environ. Microbiol.* **76**, 2445–2450 (2010).
80. Gavrish, E., Bollmann, A., Epstein, S. & Lewis, K. A trap for in situ cultivation of filamentous actinobacteria. *J. Microbiol. Methods* **72**, 257–262 (2008).
81. Bouguelia, S. *et al.* On-chip microbial culture for the specific detection of very low levels of bacteria. *Lab Chip* **13**, 4024 (2013).
- 540 82. Park, J., Kerner, A., Burns, M. A. & Lin, X. N. Microdroplet-enabled highly parallel co-cultivation of microbial communities. *PLoS One* **6**, (2011).
83. Stanley, C. E. & van der Heijden, M. G. A. Microbiome-on-a-chip: New frontiers in plant–microbiota research. *Trends Microbiol.* **25**, 610–613 (2017).
- 545 84. Scheffer, M., Hosper, S. H., Meijer, M. L., Moss, B. & Jeppesen, E. Alternative equilibria in shallow lakes. *Trends Ecol. Evol.* **8**, 275–279 (1993).
85. Schimel, J. P. & Schaeffer, S. M. Microbial control over carbon cycling in soil. *Front. Microbiol.* **3**, 1–11 (2012).
86. Raaijmakers, J. M. & Weller, D. M. Natural plant protection by 2,4-

- 550 diacetylphloroglucinol - Producing *Pseudomonas* spp. in take-all decline soils. *Mol. Plant-Microbe Interact.* **11**, 144–152 (1998).
87. Kommineni, S. *et al.* Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract. *Nature* **526**, 719–722 (2015).
88. Heijden, M. G. A. Van Der, Martin, F. M. & Sanders, I. R. Mycorrhizal ecology and evolution : the past , the present , and the future. *New Phytol.* **205**, 1406–1423 (2015).
- 555 89. van der Heijden, M. G. A. *et al.* Mycorrhizal fungal diversity determines plant biodiversity , ecosystem variability and productivity. *Nature* **396**, 69–72 (1998).
90. Shade, A. & Handelsman, J. Beyond the Venn diagram: The hunt for a core microbiome. *Environ. Microbiol.* **14**, 4–12 (2012).
- 560 91. Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Le Van, A. & Dufresne, A. The importance of the microbiome of the plant holobiont. *New Phytol.* **206**, 1196–1206 (2015).
92. Turnbaugh, P. J. *et al.* A core gut microbiom in obese and lean twins. *Nature* **457**, 480–484 (2009).
- 565 93. Yeoh, Y. K. *et al.* Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. *Nat. Commun.* **8**, 215 (2017).
94. Brown, J. H. & Heske, E. J. Control of a Desert-Grassland Transition by a Keystone Rodent Guild. *Science* (80-. ). **250**, 1705–1707 (1990).
95. Jones, C. M. *et al.* Recently identified microbial guild mediates soil N<sub>2</sub>O sink capacity. *Nat. Clim. Chang.* **4**, 801–805 (2014).
- 570 96. Lynch, M. D. J. & Neufeld, J. D. Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* **13**, 217–229 (2015).
97. Jousset, A. *et al.* Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J.* 1–10 (2017). doi:10.1038/ismej.2016.174
- 575 98. Pester, M., Bittner, N., Deevong, P., Wagner, M. & Loy, A. A ‘rare biosphere’ microorganism contributes to sulfate reduction in a peatland. *ISME J.* **4**, 1591–602 (2010).
99. Stinson, K. A. *et al.* Invasive plant suppresses the growth of native tree seedlings by disrupting belowground mutualisms. *PLoS Biol.* **4**, 727–731 (2006).
- 580 100. Manefield, M., Whiteley, A. S., Griffiths, R. I. & Bailey, M. J. RNA stable isotope probing, a novel means of linking microbial community function to phylogeny. *Appl. Environ. Microbiol.* **68**, 5367–5373 (2002).

## Figures

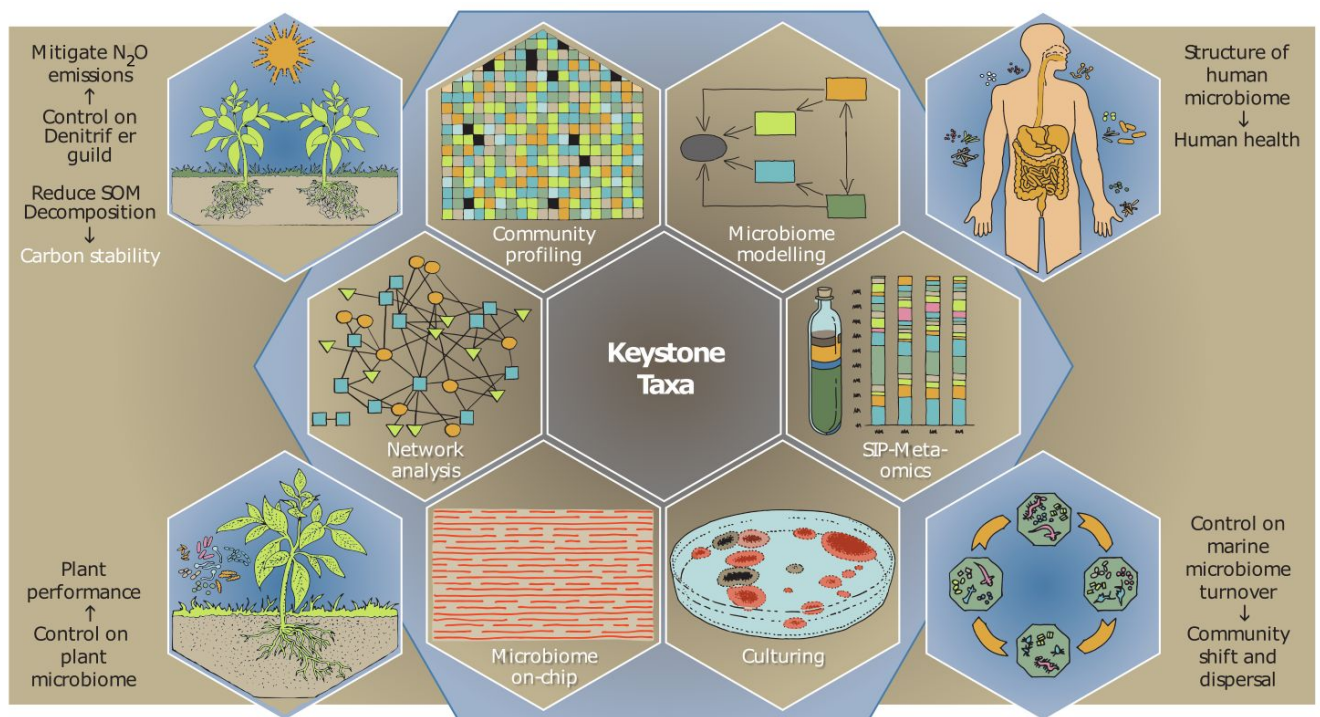


**Figure 1. A)** Modularity and keystone taxa in microbial networks. Nodes (oval shaped) represent OTUs and solid lines represent edges i.e., relationships among nodes. A network consisting of many taxa (nodes), without any highly interacting hub or keystone taxa. This network is similar to a random network that has a Poisson distribution of edges per node i.e., most nodes have similar number of edges and no highly interconnected nodes. A microbial network without any modules but with two hubs or keystone taxa. This is a scale-free network that has a power-law distribution of edges i.e., only two nodes are highly interconnected holding the network together. Highly interacting keystones positioned in two distinct clusters or modules. These keystone groups are holding the modules together. Thus, removal of such keystones may cause dramatic shift in the composition. **B)** Empirical evidence of keystone taxa in the human, plant and soil microbiomes. The numbers given in parentheses refer to the respective studies showing the keystone performance of these taxa.



**Figure 2.** Conceptual diagram showing the concept of keystone taxa in microbial communities and the factors influencing their functioning in an environment. **A)** Distinction between the modes of operation of dominant- and keystone taxa. The dominant taxa (light orange) affect microbiome functioning exclusively by virtue of sheer abundance, whereas keystone taxa (green) exert its influence irrespective of their abundance. Since the impact of dominant species on a process is primarily due to greater abundance, broadness of that process is less important. Here broadness implies that a particular process consists of many steps and are operated by diverse microbial groups. However, keystone taxa exert influence via selective modulation and might have a greater influence on narrow processes i.e., the processes that consist of single or a few steps and are operated by selected groups of microbes. **B)** Environmental and ecological factors that may determine the distribution and performance of keystone taxa in an environment. **C)** Hypothetical diagram showing various modes of functioning of keystones in an environment. Individually, keystone taxa might have greater influence on a narrow process (e.g., biological nitrogen fixation performed by highly specific microbes). On the other hand, a keystone guild comprising multiple keystone taxa within a community might also be able to influence a broad process.





**Figure 3.** Hypothetical diagram illustrating the tools for linking keystone taxa to ecosystem functioning and the research areas where keystone taxa can be employed. While network analysis can be used to statistically identify keystone taxa in microbial networks, it is important to link such taxa to ecosystem processes. With the advent of newer tools such as chip or culture-based methods, keystone taxa can be isolated from environments and cultured or co-cultured. Functional profiling of such taxa can be performed using RNA-stable isotope probing coupled with metatranscriptomics or metaproteomics. Upon functional profiling, the relative importance can be estimated through microbiome modelling. Such models involving causal relationships can be used to reveal the contribution of keystone taxa to ecosystem processes.



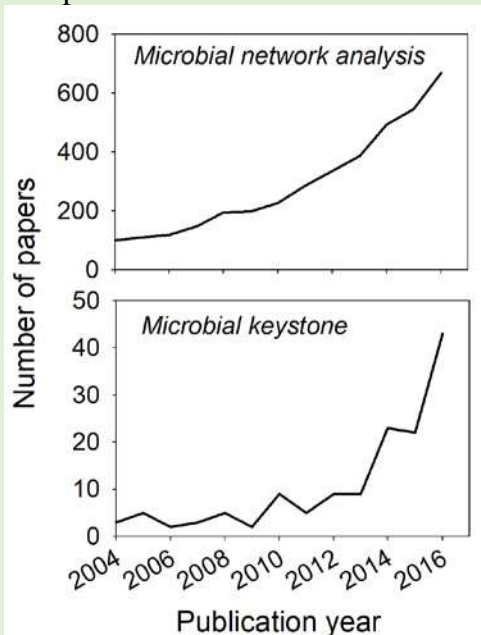
**Table 1.** Summary of studies reporting keystone taxa in different ecosystems. Within each ecosystem, studies are listed alphabetically. Members of *Rhizobiales* and *Burkholderiales* are consistently present as keystone taxa across ecosystem types, except for human microbiome. keystones across individual studies and links between keystone taxa and specific microbial ecosystem processes as reported in relevant studies are shown in **Table S1**.

<b>Ecosystem/habitat</b>	<b>Keystone taxa</b>	<b>References</b>
<i>Computational inference</i>		
Grasslands	<b><i>Burkholderiales</i></b> ; <i>Sphingobacteriales</i> <i>Clostridiales</i> ; <i>Actinomycetales</i> <i>Acidobacteria GP4</i>	Deng <i>et al.</i> 2012; Lupatini <i>et al.</i> , 2014; Zhou <i>et al.</i> , 2011
Forest/woodlands	<i>Actinomycetales</i> ; <i>Acidobacteria_Gp4</i> <b><i>Rhizobiales</i></b> ; <b><i>Burkholderiales</i></b> <i>Clostridiales</i> ; <i>Sphingobacteriales</i> <i>Rhodobacteriales</i> ; <i>Verrucomicrobia</i>	Banerjee <i>et al.</i> , 2016; Ding <i>et al.</i> , 2015; Eldridge <i>et al.</i> , 2015; Lupatini <i>et al.</i> , 2014; Ma <i>et al.</i> , 2016
Agricultural	<i>Gemmatimonas</i> ; <i>Acidobacteria_GP17</i> ; <i>Xanthomonadales</i> ; <b><i>Rhizobiales</i></b> ; <b><i>Burkholderiales</i></b> <i>Solirubrobacteriales</i> ; <i>Verrucomicrobia</i>	Jiang <i>et al.</i> , 2017; Liang <i>et al.</i> , 2016; Lupatini <i>et al.</i> , 2014; Wang <i>et al.</i> , 2017
Arctic/Antarctic	<b><i>Rhizobiales</i></b> ; <b><i>Burkholderiales</i></b> <i>Actinobacteria</i> , <i>Alphaproteobacteria</i>	Comte <i>et al.</i> , 2016; Gokul <i>et al.</i> , 2016; Hill <i>et al.</i> , 2016; Yang <i>et al.</i> , 2016
Contaminated soil	<b><i>Rhizobiales</i></b> ; <i>Nitrospira</i> ; <i>Pseudomonadales</i> ; <i>Actinobacteria</i>	Chao <i>et al.</i> , 2016; Jiao <i>et al.</i> , 2016
Plant associated microbiota	<i>Acidobacteria_GP1, GP3, GP6</i> <b><i>Rhizobiales</i></b> ; <b><i>Burkholderiales</i></b> <i>Pseudomonadales</i> ; <i>Bacterioidetes</i> <i>Frankiales</i>	Jiang <i>et al.</i> , 2017; Shi <i>et al.</i> , 2016; Yan <i>et al.</i> , 2016;
Aquatic	<i>Pelagibacter</i> ; <i>Oceanospirillales</i> <i>Flavobacteriaceae</i> ; <i>Nitrospira</i> <i>Rhodobacteriadaceae</i> <i>Alteromonadaceae</i> ; <i>Chromatium</i> ; <b><i>Rhizobiales</i></b> ; <b><i>Burkholderiales</i></b> <i>Chlorobium</i> ; <i>Verrucomicrobia</i> , <i>Chloracidobacteria</i> , <i>Chloroflexi</i> , <i>OP3</i>	Geng <i>et al.</i> , 2016; Graham <i>et al.</i> , 2017; Ji <i>et al.</i> , 2016 Musat <i>et al.</i> , 2008; Steele <i>et al.</i> , 2011; Vick-Majors <i>et al.</i> , 2014; Zhao <i>et al.</i> , 2016;
<i>Empirical evidence</i>		
Agricultural*	<i>Gemmatimonas</i> ; <i>Acidobacteria</i>	Banerjee <i>et al.</i> , 2016b; Li <i>et al.</i> , 2017
Phyllosphere	<i>Albugo</i> , <i>Dioszegia</i>	Agler <i>et al.</i> , 2016
Human microbiome		
Oral microbiome	<i>Porphyromonas gingivitis</i> <i>Streptococcus gordonii</i>	Hajishengallis <i>et al.</i> , 2011 Hajishengallis <i>et al.</i> , 2012
Gut microbiome	<i>Helicobacter pylori</i> ; <i>Methanobrevibacter smithii</i> ; <i>Actinobacteria</i> ; <i>Bacterioides fragilis</i> ; <i>Bacterioides stercosis</i> ; <i>Bacterioidetes thetaiotaomicron</i> ; <i>Ruminococcus bromii</i> ; <i>Klebsiella pneumoniae</i> ; <i>Proteus mirabilis</i>	Curtis <i>et al.</i> , 2014; Fisher and Mehta, 2014; Garrett <i>et al.</i> , 2010; Maldonado-Contreras <i>et al.</i> , 2011; Shetty <i>et al.</i> , 2017; Trosvik and

\*Keystone taxa were initially identified using network based scores and linked to organic matter decomposition

### **BOX 1. Microbial co-occurrence and network analysis**

The number of papers on Web of Science database reporting microbial network analysis and keystone taxa is increasing exponentially as shown in the figure below. Papers were searched until 2016 using the keywords *microbial network analysis* and *microbial keystone*. A wide range of methods and algorithms are available to construct microbial networks with each having its own usefulness. Starting with basic Pearson or Spearman rank correlation based approaches<sup>7,14</sup>, microbial networks quickly evolved to incorporate more robust methods. For example, maximal information coefficient (MIC) developed by Reshef et al (2011), relies on equitability and generality of relationships, and can yield a variety of linear and nonlinear associations among microbes that can be interpreted similar to coefficient of determination i.e.  $r^2$ . Local similarity



analysis (LSA) that detects change in abundance of operational taxonomic units (OTUs) over time in an environment, is particularly useful for analyzing microbial temporal variability data<sup>16</sup>, whereas SparCC is especially suited for compositionally diverse microbial data<sup>17</sup>. On the other hand, ensemble approach (CoNet) can use multiple measures (similarity, correlation, mutual information) with generalized boosted linear model to generate comprehensive networks<sup>18</sup>. Recently, Weiss et al (2016) found LSA, MIC and SparCC are well-suited for both count and compositional data and are less sensitive to distribution

shape of data, whereas CoNet ensemble approach performs better for data with high scatteredness or sparsity. However, they also noted that different approaches may yield different results and significance levels for the same dataset, and while the scores obtained from thousands of pairwise correlations are typically corrected for Type I error using Bonferroni or false discovery rate, extremely rare OTUs or OTUs with large number of zeros should be avoided for network construction.